DOCUMENT RESUME

ED 219 231 SE 038 256

AUTHOR Sweitzer, Gary L.
TITLE A Meta-Analysis of Research on Preservice and Inservice Science Teacher Education Practices Designed to Produce Outcomes Associated with Inquiry Strategy.
PUB DATE Apr 82
GRANT SED-80-12310
NOTE 38p.; Paper presented at the Annual Meeting of the National Association for Research in Science Teaching (55th, Chicago, IL, April 5-8, 1982). This research was conducted as part of a larger meta-analysis project initiated at the University of Colorado.

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Elementary School Science; Elementary Secondary Education; Higher Education; *Inquiry; *Inservice Teacher Education; *Preservice Teacher Education; Research Methodology; Science Education; *Science Instruction; Science Teachers; Secondary School Science; Student Characteristics; Teacher Behavior; Teacher Characteristics; *Teaching Methods
IDENTIFIERS *Meta Analysis; *Science Education Research

ABSTRACT

The discrepancy between educators' expectations for inquiry behavior and the actual status of such behavior (teachers feeling more comfortable teaching facts and feeling ill-prepared to guide students in inquiry learning) prompted a quantitative assessment of the existing research (reported between 1965 and 1980 in doctoral dissertations and journals) on training outcomes associated with inquiry teaching behaviors and the techniques and procedures used to obtain them. Meta-analysis accomplished this assessment and studies (N=68) were selected that had at least one outcome associated with inquiry strategy (knowledge of science processes, inquiry instructional strategy, indirect verbal behavior, accepting interpersonal behaviors, increased wait-time questioning behavior, higher cognitive level questioning, and discovery instructional strategy). Relevant variables were identified and coded into six major categories: 1) study form/design characteristics; 2) teacher/teacher trainee characteristics; 3) student characteristics; 4) treatment characteristics; 5) outcome characteristics; and 6) effect size characteristics (standardization of mean differences between treatment and control groups). Results among others indicated the teacher outcome most frequently measured was knowledge of science processes followed by indirect verbal behavior. (Author/JN)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED219231

A Meta-Analysis of Research on Preservice and Inservice Science Teacher Education Practices Designed to Produce Outcomes Associated with Inquiry Strategy

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
✓ This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

Gary L. Sweitzer
Ohio State University
and
Upper Arlington City Schools
Columbus, Ohio

National Association for Research in Science Teaching
55th Annual Meeting
April 1982

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Gary L. Sweitzer
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

This research was conducted as part of a larger meta-analysis project initiated at the University of Colorado under NSF grant SED80-12310, Ronald D. Anderson, Director.

SE 031 206

INTRODUCTION

The meta-analysis steering committee identified "What are the effects of different preservice and inservice science teacher training approaches" as one facet of science education having potential for meta-analysis. With this question in mind an initial identification of relevant studies was accomplished through a title search of the dissertations listed in the Science Education Dissertation Bibliography compiled by Dr. Stanley Helgeson. This search identified approximately 215 dissertations produced between 1950 and 1977 which addressed the proposed question. This volume of research was considered beyond the scope of this project and a rationale for limiting the question was investigated. The basis for this rationale centered on the science course improvement projects of the sixties and seventies. These projects sought to improve the science education of young people by placing a greater emphasis upon rationale thinking as a course outcome, using the discipline as a criterion for the selection of instructional materials, organizing the curriculum with both a concept and inquiry sequence, and shifting more responsibility for learning to the student. The essence of this curriculum reform mandated the teaching of science in an inquiry style of teaching that was investigative and student centered with inquiry skills as outcomes of instruction.

Once we have identified a desired classroom behavior, here an inquiry style of teaching, the next step is determining teacher training procedures that can produce that behavior.

A review of the training research by Balzer, Evans, and Blosser indicated the following:

1) An examination of 71 investigations revealed that providing teachers with training in systematic observation of classroom behavior was an effective means of changing classroom behavior. This was true of the science education and non-science education investigations.

2) Systematic observations was used in 29 studies to determine the influence of inservice training in one of the "New Science" programs on science classroom behavior. A tally of the reported results revealed that inservice training in one or more of the course content improvement projects and/or use of the materials was a fairly effective way of influencing selected classroom behaviors.

3) Microteaching, either by itself or in combination with other instructional sequences, was reported to have been effective in changing selected aspects of classroom behavior in 7 of 10 studies involving science classrooms. Selected aspects of behavior referred to specific teaching skills which were generally predetermined, operationally defined and practiced in a microteaching format.

4) Feedback following teaching was used in an attempt to alter classroom behavior in 38 investigations. Feedback was divided into the following 4 types: student data from systematic observation of classroom behavior, videotaped and/or audiotaped classes, and supervisory conferences. The results revealed that individually the subclasses of feedback were ineffective

or inconsistent change agents, but that supervisory conference and videotaped feedback in combination brought about changes in selected aspects of classroom behavior. (1) While this is only a sample of the research reviewed by Balzer, Evans, and Blosser, it would seem that techniques have been developed that accomplish teacher classroom behavior change.

However, Welch et al in "The Role of Inquiry in Science Education: Analysis and Recommendations" states "There appears to be a discrepancy between existing general statements about the importance of inquiry and the attention given it in practice. Although teachers made positive statements about the value of inquiry, they often felt more responsibility for teaching facts, things which show up on tests, and structure of the work ethic. A major problem in promoting inquiry was encountered in the preparation of science teachers. Many teachers are ill-prepared, in their own eyes and in the eyes of others to guide students in inquiry learning and over one-third feel they receive inadequate support for such teaching." (16)

This discrepancy between educators expectations for inquiry behavior and the actual status warranted a quantitative assessment of the existing research on training outcomes associated with teaching inquiry behaviors and the techniques and procedures used to obtain them. Meta-analysis accomplished this quantitative assessment and studies were selected that had at least one outcome associated with inquiry strategy.

The basis for a definition of inquiry strategy must be set in historical perspective. While the curriculum reform movement of the sixties and seventies brought renewed consideration of inquiry strategy Hurd indicated that the Progressive Education Association in its publication Science in General Education (1938) described a program emphasizing the inquiry and social aspects of science. In 1945, a Harvard University Committee reporting in General Education in a Free Society recommended the teaching of high school science using broad integrative elements and scientific modes of inquiry set within cultural, historical, and philosophical context. J. Darrell Barnard in a 1968 review of pre-1960 contributions to science education states "Failure to find such terms as 'discovery' and 'process' in the earlier literature on research in science teaching does not mean that their referents have not been investigated. In fact, they have, and such concepts as 'scientific thinking', 'problem solving', 'inductive method' and the like have much in common with the equally vague concept currently referred to as 'discovery and process'. This situation is only one example of the semantic problems which make interpretation, evaluation, and replication of research in science teaching so extremely difficult". (5)

Kyle addressed this semantics problem in an effort to make the distinction between inquiry in general and scientific inquiry and summarized several different activities and strategies associated with inquiry.

Kyle asserted that there was little agreement as to what constitutes scientific inquiry and indicated that scientific inquiry should not be construed as synonymous with investigative experimental or discovery methods of science teaching, self-instructional learning techniques, or open-ended learning techniques. Kyle stated "the ability to scientifically inquire is the personal, internalized ability of an individual to synthesize the knowledge which has been obtained through the learning of basic process skills and competencies, that enables a person to rationally inquire and solve problems by means of unrestrained inductive thinking". (8)

Welch et al offered further clarification on the semantics of inquiry considering "inquiry to be a general process by which human beings seek information or understanding. Broadly conceived, inquiry is a way of thought. Scientific inquiry, a subset of general inquiry, is concerned with the natural world and guided by certain beliefs and assumptions". Welch et al divided inquiry into three main themes: 1) general inquiry processes, 2) science process skills, 3) nature of scientific inquiry. General inquiry processes include strategies such as problem-solving, use of evidence, logical and analytical reasoning, clarification of values, decision-making, and safeguards and customs of inquiry. Science process skills include the usual range of science processes, such as observing, measuring, interpreting data, etc. The nature of scientific inquiry is affected by the structure of scientific knowledge and by

assumptions about the natural world such as causality and non-capriciousness. Inquiry strategy within the context of this paper addressed those teacher behaviors that facilitate student acquisition of concepts, processes, and skills through active involvement with general inquiry strategies. It incorporates aspects of the investigative and discovery phases of science and affords opportunities for the students to test and refine concept meanings. Through this type of learning and the acquisition and synthesis of scientific knowledge and processes the ability to perform scientific inquiry becomes possible. A teacher equipped to conduct inquiry would possess questioning skills that are divergent, have a knowledge of science processes and have the capability of conducting a student-centered inductive approach. Specific outcome criteria which appeared in the research reviewed for this analysis and were judged as falling within the confines of this definition included the following: knowledge of science processes, inquiry instructional strategy, indirect verbal behavior, accepting interpersonal behaviors, increased wait-time questioning behavior, higher cognitive level questioning behavior, and discovery instructional strategy.

METHOD

Research dated between 1965 and 1980 was considered for this study. This paralled the implementation period for the modern science curriculum projects, and most of the research during this period was well documented by three ERIC compilations on science educational research.

The Colorado staff reviewed the titles of dissertations listed in the Science Education Dissertation Bibliography compiled by Dr. Stanley Helgeson and identified dissertations related to in-service and pre-service teacher education. A review of these dissertations resulted in the development of the rationale for limiting the study to research which involved outcomes associated with inquiry strategy. This review and work by Munby suggested a further limitation when Munby concluded that "on the basis of his analysis there are grounds for viewing research on the affective outcomes of science education with misgiving, simply because there seems little to be said of the instruments as to enlist our confidence in their use". (11) Furthermore, if we are attempting to change attitude with an eye toward teacher practice, a review of research on changing the attitude of student teachers by Morrisey (10) indicated that the lack of change in the teaching of elementary science in the schools indicates that something more than just immediate attitude change must be considered. Therefore, studies involving only affective outcomes were not included in the meta-analysis.

Journal articles were identified applying the same criteria as for dissertations by scanning the table of contents of the Journal of Research in Science Teaching and Science Education for the years 1965 to 1980. Research from sources other than dissertations and journals was identified through a search of the three ERIC compilations of abstracts from Resources in Education (1966-1972, 1973-1975, 1976-1977). Most of the materials identified in the compilations duplicated materials found elsewhere. The reasons for rejecting studies for the meta-analysis included the following: data for the calculation of effect size were incomplete, measured outcomes were only in the affective domain, the native language of the subjects was other than English, the study was produced outside the 1965 to 1980 range, the outcome variables were not associated with inquiry teaching strategy.

Relevant variables were identified and coded according to the following six major categories: study form and design characteristics, teacher/teacher trainee characteristics, student characteristics, treatment characteristics, outcome characteristics, and effect size calculation characteristics. These categories and six variables for the management of the data deck resulted in the delineation of 114 variables.

Effect size calculations were performed using the most straightforward method possible with the data presented in each study.

The most straightforward method available and the one used in 64% of the effect size calculations involved standardization of the mean differences between treatment and control groups. This had the advantage of allotting equal effect sizes to equal means. (4) Other effect size calculation procedures are addressed in the section on effect size calculation characteristics.

ANALYSIS OF DATA
INTRODUCTION AND STUDENT OUTCOMES

Sixty-eight studies were coded resulting in 177 effect size calculations. Outcomes were measured on teachers, on students, and on students about teachers, While many advocate measuring teacher behavior by evaluating student performance this occurred in a very small number of cases. One hundred and fifty-four effect sizes were outcomes measured on teachers while only nineteen were outcomes measured on students and four on students about teachers. Because of the small sample size no analysis of the effect sizes related to outcomes measured on teachers by students was performed.

The nineteen effect sizes related to outcomes measured on students produced a mean effect size of .436 and a standard deviation of .670 when broker down across all variables. The mean reliability of the measurement instruments was .818 with a standard deviation of .063. The outcome most frequently measured was the knowledge of science processes occurring in 47.4% of the cases.

ANALYSIS OF DATA
TEACHER OUTCOMES

The one hundred and fifty-four effect sizes related to outcomes on teachers produced a mean effect size of .848 with a standard deviation of 1.297 when broken down across all variables. The size of the teacher samples ranged from 9 to 299 with a mean of 60.4 and a standard deviation of 45.2. The number of teachers assigned to each treatment ranged from 5 to 129 with a mean of 26.8 and a standard deviation of 17.6. Test reliability was reported for fifty percent of the outcome measure instruments yielding a mean of .81 and a standard deviation of .09. Duration of treatment (with seven missing cases) ranged from less than a day to one year. The mean was 70.0 days with a standard deviation of 71.4.

One effect size determination was considered as a far outlier having a value three times greater than the next closest effect size and being approximately ten standard deviations above the mean. With this value discarded the mean effect size across all variables becomes .77 with a standard deviation of .86.

Descriptive reporting of teacher and/or student characteristics that might affect outcome measures was sporadic and occurred with frequencies too low to support analysis. Therefore the analysis was divided into the following three major categories: variables associated with methodological aspects of the studies, variables associated with study outcomes and treatments, and

variables associated with effect size calculations.

Table I presents the means, standard deviations, and sample sizes of selected methodological variables broken down across outcomes measured on teachers. Table II presents the Pearson correlation coefficients and significance level between selected methodological variables and between selected methodological variables and effect size. In situations where the variable categories did not reflect an ordinal relationship the categories were coded dichotomously for correlational purposes.

Table I presents nine methodological variables with the categories that had sufficient sample size to warrant discussion. The form of publication variable indicated the source of the information used for coding purposes. An attempt was made to locate the primary source whenever possible and if a study was presented as both a journal article and a dissertation, the dissertation was used as the source. The mean effect size calculated from journal articles was 1.01 while those calculated from dissertations was .59. There was a significant positive correlation ($p = 0.004$) between journals as a source and effect size and a significant negative correlation ($p = .005$) between dissertations as a source and effect size.

The type of study variable was coded using the guidelines established by Campbell and Stanley in their chapter on experimental designs in N. L. Gage's First Handbook of Research on Teaching.

Table I
Teacher Outcome Effect Sizes
Across Study Methodological Variables

Variable	Mean Effect Size	Standard Deviation	N
Form of publication			
Journal	1.01	.98	61
Dissertation	.59	.77	85
Other unpublished	.75	.24	7
Type of study			
Pre-experimental	.90	.73	22
Quasi-experimental	.78	.83	69
Experimental	.67	.92	60
Assignment of teachers to treatments			
Random	.67	.91	61
Self-selected	.57	.78	17
Intact groups	.88	.86	56
Teacher unit of analysis			
Individual	.77	.88	145
Classroom or group	.75	1.00	4
Rated internal validity			
Low	.55	.55	25
Medium	.75	.82	55
High	.82	.98	68
Design Rating			
Low	.57	.80	41
Medium	.68	.61	38
High	.92	.99	72
Outcome instrument type			
Published, national stand.	.35	.58	16
Ad hoc, for that study	1.12	.94	59
Other	.60	.76	77
Measurement method			
Multiple choice	.48	.61	31
Likert	.50	.44	24
Observation	.84	.93	49
Other	1.14	1.07	35

Variable	Mean Effect Size	Standard Deviation	N
Time of measurement			
After treatment	.62	.76	60
Pre-post	.90	.96	80
Other	.74	.59	9

Table II
Correlations Among Selected Methodological
Variables and Between Methodological Variables and Effect Size

Variable A	Variable B	r	p	N
Form of pub. journal	effect size	0.2308	0.004	61
Form of pub. journal	extent of treat	0.2140	0.008	61
Form of pub. journal	# of teachers assigned	-0.0390	0.634	61
Form of pub. journal	# of teachers analyzed	-0.0042	0.959	61
Form of pub. journal	reported sig.	-0.2153	0.010	61
Type of study	effect size	-0.0945	0.248	151
Rated internal validity	effect size	0.1027	0.214	148
Design rating	effect size	0.1719	0.035	151
Outcome instrument, pub. national standardized	effect size	-0.1670	0.039	16
Outcome instr. ad hoc	effect size	0.3207	0.000	59
Outcome instr. other	effect size	-0.2017	0.012	77
Measurement method multiple choice	effect size	-0.1674	0.039	31
Measurement method other	effect size	0.2363	0.003	35
# of teachers assigned	effect size	-0.2063	0.011	152
# of teachers analyzed	effect size	-0.1748	0.031	153
# of teachers on outcome measure	effect size	-0.1732	0.035	149

Pre-experimental designs consisted of one-group pretest-posttest designs and static-groups comparison designs. Experimental designs involved pretest-posttest control groups designs and posttest-only control group designs with random assignment of subjects to treatment and control groups. Quasi-experimental designs involved the same designs as the experimental groups, but without random assignment of subjects. The mean effect size ranged from .90 for pre-experimental studies to .67 for experimental studies. The correlation coefficient was not significant at the .05 level.

The assignment of teachers to treatments was coded as random, matched, self-selected, intact groups, representative sample, and other. The teacher unit of analysis variable indicated the unit that was used in establishing the degrees of freedom for the determination of significance level. Random assignment and intact group assignment together accounted for 76% of the effect sizes produced. Random assignment studies had an average effect size of .67 and intact groups produced an effect size mean of .88. The teacher unit of analysis used the number of individuals involved to determine degrees of freedom in 95% of the cases.

The internal validity variable addressed the assignment of individuals to treatments, and the percent mortality among treatment and control groups. Studies were rated high in internal validity if group equivalence was established through random assignment or other procedures and subject mortality was less than 15 percent.

Studies were coded as medium if a randomization design failed but maintained low mortality, if randomization existed but mortality was high or nonequivalent, and if random procedures were used in the selection of intact groups and mortality was low. Studies were coded as low where intact convenience samples were used and/or where the existence of factors confounding the equivalence of the subject groups was apparent. Studies rated low in validity produced effect sizes with a mean of .55, medium produced effect sizes with a mean of .75, and high produced effect sizes with a mean of .82. The internal validity produced a positive correlation with effect size, but it was not significant at the .05 level.

The design rating variable addressed the apparent degree of control of the confounding variables by the experimental procedure. Studies were rated as high if the design established control to the extent that posttreatment differences could be attributed to treatment effects. The study was rated as medium if the design indicated lack of control of a variable that probably contributed to some outcome differences. The study was rated as low if the failure to control a given variable obviously contributed to outcome differences. Studies rated as low produced an average effect size of .80, medium .61, and high .99. A positive correlation coefficient with a significance level of .035 indicated that larger effect sizes were associated with higher design ratings.

The outcome instrument type variable categorized instruments as published, as ad hoc being developed for that particular study, as departmental or locally standardized, as classroom

based or teacher developed, or other. Most of the tests in the other category were developed ad hoc for another study and then used in existing form or with modifications for the study being coded.

The measurement method variable included the categories of multiple choice, semantic differential, Likert, questionnaire, observation interview, Q-sort, and other. The other category included essay tests, short answer tests and instruments that involved producing actual examples of the desired skill (eg. question writing).

The effect sizes that resulted from ad hoc instruments produced a mean effect size of 1.12. Other instruments produced an effect size of .60 and published instruments a mean of .35. The ad hoc instruments had a significant positive correlation with effect size ($p < 0.001$) and the published and other categories had a significant negative correlation with effect size ($p < .05$). Measurement method categories produced the following results for mean effect size: multiple choice .48, Likert .50, observation .84, and other 1.14. Multiple choice methods correlated negatively with effect size with a significance of .039. The "other" category correlated positively with effect size with a significance of .003. The remaining categories did not correlate significantly at the .05 level.

The time of measurement variable included the following categories: before treatment, after treatment, pre-post,

delayed, and other. The other category was used when different instruments on the same outcome were averaged to determine one effect size. The pre-post measurement produced the largest mean effect size, .90. The "other" category had a mean effect size of .74 and the after treatment category .62.

The number of teachers assigned to the study, the total number analyzed, and the number measured on each outcome instrument correlated negatively with effect size with a significance level of p .05. The journal category of the form of publication variable did not correlated significantly at the .05 level with any of the variables related to sample size, but the journal category did correlate significantly with the reported significance variable and the extent of treatment variable. The extent of treatment variable addressed the scope of the treatment with a multi-grade treatment, a program, or an on-going institute being at the broad end of the scale and a specific training technique being at the narrow end of the scale. These correlations indicated that studies taken from journals had low values for p and that these studies addressed treatment types of narrow scope.

Table III presents the average effect size, standard deviations and correlation coefficients for two of the variables related to effect size calculations. Means for the calculation of effect size were unadjusted post-test means in 83 cases and provided a mean effect size of .62 with a standard deviation of .74. Means were a result of pre-post differences in 47 cases with a mean effect size of 1.00 and a standard deviation of 1.09. The

Table III
Teacher Outcome Effect Sizes
Across Effect Size Calculation Variables

Variable	Mean Effect Size	Standard Deviation	N
Source of means unadjusted post-test	.62	.74	83
Source of means pre-post differences	1.00	1.09	47
Source of means other	.88	.75	19
Calculated directly from reported values or raw data	.72	.90	96
Calculated with direct estimates (ANOVA,t,F)	.84	.87	34
Calculated from reported probability levels	.71	.15	5
Calculated using pre-test data as a control group	1.02	.78	15

Correlation with effect size

Variable	r	p	N
Source of means unadjusted post-test	-0.1865	0.021	83
Source of means pre-post dif.	0.1745	0.031	19
Calculated directly from reported values or raw data	-0.0694	0.394	96
Calculated with direct estimates (ANOVA,t,F)	0.0420	0.606	34
Calculated from reported probability levels	-0.0131	0.872	5
Calculated using pre-test data as a control group	0.0948	0.244	15

"other" category involved pre-experimental studies wherein pre-test data was used to generate a control group mean. This category provided a mean effect size of .88 and a standard deviation of .75. The unadjusted post-test source of means had a negative correlation with effect size significant at the .05 level (p=.021). The source of means in the pre-post category had a positive correlation with effect size with a probability level of .031.

The methods used for calculating the effect size included the following: calculating directly from reported means and variances or from raw data, calculating with direct estimates of the variance from ANOVA, t, and F values, calculating from non-parametric tests by substituting the t value that had the equivalent level of significance, calculating by using variances from other independent sources, calculating using reported probability levels, and calculating using pre-test data as a control group. Using pre-test data as a control group produced the highest mean effect size 1.01 with a standard deviation of .78. The means and standard deviations for using directly reported means and variances and direct estimates were close having mean effect values of .72 and .84 respectively. None of the calculation method variables showed a significant correlation with effect size at the .05 level.

Table IV presents the effect sizes of teacher outcomes broken down across various treatment variables and Table V presents correlations among selected treatment variables and between treatment variables and effect size. The time of

Table IV
Teacher Outcome Effect Sizes
Across Treatment Variables

Variable	Mean Effect Size	Standard Deviation	N
Time of treatment			
Pre-service	.78	.90	122
In-service	.72	.74	31
Site of treatment			
Field based, site of employment	.74	.86	5
Field based, not the site of employment	.77	.60	20
University based	.77	.88	112
Extent of treatment			
Multi-grade or level, eg. program or ongoing institute	.45	.45	12
One grade or level, eg. course or workshop	.75	.78	69
Training technique	.84	.98	72
Treatment geared to grade level			
Elementary school	.76	.86	123
Secondary	.39	.32	8
General	1.24	.97	15
Number of variables used to describe each treatment			
1	.67	.57	42
2	.65	.89	64
3	.73	.62	31
4	1.25	1.03	14
Treatment type (N greater than 7)			
Field based program	.35	.40	8
Workshop	.73	.75	16
Methods course	.79	.94	22
Science course designed for teachers	.97	.70	9
Unit of study	1.38	1.29	22
General instructional exposure	.79	1.21	35
Inquiry instructional exposure	.63	.63	9

Variable	Mean Effect Size	Standard Deviation	N
Treatment type cont'd			
Instructional exposure concrete mode	.75	.75	20
Instructional exposure mixed	.44	.86	12
Source of structure student self-directed	.04	.46	8
Source of structure student interacting with materials and/or the teacher	.70	1.01	8
Instructional exposure locus of control, student self-directed	.81	.88	44
Instructional strategy feedback	.67	.91	10
Modeling strategy	1.56	1.20	14
Behavior coding training or strategy analysis	1.37	.87	8
Questioning analysis	1.38	1.65	8
Video technology	1.82	1.44	9
Programmed material or a-t	.99	.76	17
Outcome criteria			
Knowledge of science content	.52	.79	7
Knowledge of science process	1.08	1.03	33
Indirect verbal behavior	.72	.82	18
Questioning level	.72	1.18	13
Discovery strategy	.70	.53	7
Attitude toward science	.39	.29	10
Questions, process directed	1.45	.60	9

Table V
Correlations Among Selected Treatment Variables
and Between Treatment Variables and Effect Size

Variable A	Variable B	r	p	N
# of variables describing treatment	effect size	0.3123	0.000	153
treatment units of study	effect size	0.2884	0.000	22
source of structure student self-directed	effect size	-0.2003	0.013	8
questioning analysis	effect size	0.1664	0.040	8
modeling strategy	effect size	0.2928	0.000	14
behavior coding training or strategy analysis	effect size	0.1637	0.043	8
video technology	effect size	0.3039	0.000	9
outcome science processes	effect size	0.1871	0.021	33
outcome questions process directed	effect size	0.1962	0.015	9

treatment was categorized as pre-service and in-service and these two groups produced mean effect sizes of .78 and .72 respectively.

The site of treatment variable categorized field based treatments versus university based treatments. The field based treatments category was further divided into treatments that occurred in the schools in which the teachers were employed versus treatments in schools where the individuals were not employed. Treatments were predominately university based accounting for 77% of the reported effect sizes. The mean effect sizes for the three groups were very close in value ranging from .74 to .77.

The extent of treatment variable ranged from a broad scope addressing multi-grade or program treatments to treatments that focused on a particular training technique. The multi-grade or program level produced a mean effect size of .45, the one grade or level variable produced a mean effect size of .75, and the training technique level produced a mean effect size of .84.

The treatment geared to grade level variable categorized the target population were the treatment outcomes were to be applied. The elementary level group accounted for 81% of the effect sizes coded and produced a mean effect size of .76 with a standard deviation of .86. Those treatments categorized as secondary had a mean effect size of .39 and those categorized as general had a value of 1.24.

In many instances more than one treatment variable was used to classify a treatment. Those treatments that were described using one variable produced a mean effect size of .67, two variables .89, three variables .73, and four variables 1.25. A positive correlation between this variable and effect size occurred at a probability level $< .001$.

Treatment type variables were divided into a number of categories. Some treatments were categorized in terms of the treatment organizational pattern and included the following: competency based programs, field based programs, on-going institutes, summer institutes, workshops, methods courses, science courses, science courses designed for teachers, mini-courses, and units of study. The field based program category resulted in a mean effect size of .35. The workshop level produced a mean effect size of .73, methods courses .79, science courses designed for teachers .97 and specific units of study 1.38. The units of study category had a positive correlation with effect size at a probability level $< .001$.

A second category involved classification of the type of instruction involved in a treatment. Instruction was broken into the following five categories: strategy, mode, interaction, source of structure, focus of control. The strategy concept addressed the classification of instructional approach. If the treatment involved instruction versus no instruction with no further delineation of approach it was categorized as

general. Other approaches were termed traditional, inquiry, and discovery. The classification as inquiry or discovery was made from the language used in the study coded even though the terms were considered as synonomous in this report. The general instructional category produced a mean effect size of .79 and the inquiry category a mean effect size of .63.

The mode of instruction categorized the approaches as predominately verbal, as predominately concrete indicating a high level of student involvement with manipulative exercises, or as mixed involving both the verbal and the concrete. The concrete mode produced a mean effect size of .75 and the mixed mode produced a mean effect size of .44.

The interaction category classified the verbal component of instruction as direct, mixed, or indirect. While this type of variable is often identified as a desireable outcome, it was not identified as a treatment in any of the studies.

The source of structure concept addressed the source of instructional objectives, content, and/or method used in the treatment. The categories included student self-directed, student interacting with materials/and or the teacher, the teacher as the source of structure, and criterion referenced sources. The source of structure involving student self-direction produced a very low average effect size, .04. The eight cases wherein this occurred produced a negative correlation with effect size

at a probability level of 0.013. The structure that involved interaction of the student with materials and/or the teacher produced a mean effect size of .70.

The focus of control concept addressed the approach used in meeting the objectives etc. set forth in the structure component. The categories involved student self-directed, student and teacher working together, teacher directed, and a mix of part student and part teacher directed. The student self-directed category produced a mean effect size of .81.

The training technique concept addressed educational practices usually employed within the confines of a course or workshop. This concept involved the following categories: interaction analysis feedback, instructional strategy feedback, wait-time analysis, questioning analysis, micro-teaching peers, micro-teaching students, modeling strategy, behavior coding training or strategy analysis, interview training, question construction activities, or the use of persuasive communications that present the benefits of a particular approach. Instructional strategy feedback produced a mean effect size of .67, modeling strategy 1.56, behavior coding training analysis 1.37, and questioning analysis 1.38. Questioning analysis, modeling strategy, and behavior coding training or strategy analysis correlated positively with effect size with a probability level less than .05.

The technology concept addressed the use of audio

technology, video technology, computer technology, programmed material and auto-tutorial methods, and print materials as treatment variables. Treatments using video technology produced a mean effect size of 1.82 and programmed material or auto-tutorial a mean effect size of .99. Video technology used in nine cases produced a positive correlation with effect size at a probability level less than .001.

Outcome criteria were classified into the following four categories: criteria related to knowledge and intellectual processes, criteria related to classroom teacher behaviors, criteria related to affective outcomes, and criteria related to curriculum characteristics and implementation, and a miscellaneous category. Only criteria with a cell size of seven or more were analyzed.

In the knowledge and intellectual processes category knowledge of science processes occurred in 33 cases and produced a mean effect size of 1.08. Classroom teacher behaviors produced a mean effect size of .72 for indirect verbal behavior, .72 for questioning level, and .70 for discovery strategy. In the affective category attitude toward science had a mean effect size of .39 while in the miscellaneous category qyestions that were process directed had a mean effect size of 1.45. Only two criteria with an N greater than seven correlated significantly with effect size at the .05 level; thse were the science process outcomes and the outcome involving questions that were process

directed.

Since the outcome criteria were varied and the cell sizes were small effect sizes were analyzed by blocking on the outcome classifications of knowledge, teacher classroom behavior, affective outcomes, and curriculum related outcomes. The knowledge category produced 52 effect sizes with a mean of .78 and a standard deviation of .93. The teacher classroom behavior category resulted in 49 effect size calculations with a mean of .71 and a standard deviation of .81. The affective outcome category had 27 effect size calculations with a mean effect size of .49 and a standard deviation of .46. The curriculum category only produced three effect size calculations.

The only treatment with a significant positive correlation at the .05 level with effect size in the knowledge category was the unit of study treatment. This treatment occurred in nine cases and had a correlation of .39. The treatment wherein the student is self-directed in producing the course structure had a significant ($p < .05$) negative correlation with effect size and occurred in six cases.

In the teacher classroom behavior block those treatments with a significant ($p < .05$) positive correlation with effect size and a cell size greater than 4 included the unit of study treatment, questioning analysis training, training using a modeling strategy, and behavior coding training or strategy analysis. A mixed instructional exposure treatment had a

significant negative correlation with effect size and occurred in 8 cases.

Conclusion

In conducting a meta-analysis the source of data should be considered. In this study journals as a source of data correlated positively ($p < .05$) with the magnitude of effect size and the average journal effect size was almost twice the average value obtained from dissertations. A meta-analysis based solely on journals as a source has the potential for establishing effect size data higher than what might be expected if primary sources of data were used.

The type of study and the rated internal validity of the study did not correlate significantly with effect size. The variable addressing the degree to which the experimental design controlled possible confounding variables correlated positively with effect size indicating that the better controlled studies produced higher effect sizes.

Experimental designs that used outcome instruments specifically designed or adapted for the study in question were positively correlated (sig. at .05) with effect size magnitude. This might be a result of investigator bias in the design and use of the instrument or the instrument might be better designed to detect outcome differences for the study in question. Measurement instruments that used a multiple choice format were significantly negatively correlated with effect size indicating that these instruments were associated with smaller effect size measures.

The use of unadjusted post-test means had a significant negative correlation with effect size indicating that these

were associated with smaller effect size measures. The use of means that are a result of pre-post differences had a significant positive correlation with effect size magnitude.

The method used to calculate the effect size did not correlate significantly ($p < .05$) with effect size. The small value of the correlations between the method of calculation and effect size affirms that the method of calculating the effect size did not contribute much to the variance of the sample.

While many advocate measuring teacher behavior by student performance this practice occurred infrequently in this sample, and descriptive reporting of teacher and/or student characteristics that might affect measured outcomes was sporadic and limited. Furthermore, while many of the outcomes associated with inquiry involve the higher levels of cognitive development, an overwhelming majority of the treatments in this sample (80%) were geared toward the elementary teaching level.

The teacher outcome most frequently measured was knowledge of science processes occurring in 21% of the effect size calculations with a mean of 1.08 and a standard deviation of 1.03. The second most commonly measured teacher outcome was indirect verbal behavior occurring in 11.7% of the cases with a mean of .72 and a standard deviation of .82. The science process outcome and the questions process directed

outcome had a significant positive correlation with effect size ($p < .05$) indicating that treatments produced the largest differences in these areas.

The treatment description variable most commonly identified was that the focus of control was student self-directed. This occurred in forty-four cases and was associated with a mean effect size of .82 and a standard deviation of .88. The second most commonly identified treatment was general instructional exposure. This occurred in thirty-five cases and produced a mean effect size of .79 with a standard deviation of 1.2 The following five treatment descriptions had a positive correlation with effect size significant at the .05 level: units of study, questioning analysis training, modeling strategy, behavior coding training or strategy analysis, and video technology. Instructional procedures using these treatments produced the largest differences in the outcomes in question.

REFERENCES

1. Balzer, A. L.; Evans, T. P.; and Blosser, P. E. A Review of Research on Teacher Behavior Relating to Science Education. AETS and ERIC Information Analysis Center for Science, Mathematics, and Environmental Education, December 1973.

2. Baumgart, Neil, and Low, Brien. "Reporting Practical and Statistical Significance in Science Teaching Research." Journal of Research in Science Teaching. Vol. 17, No. 3. 1980, pp. 269-274.

3. Craven, Gene F. "Preservice Science Teacher Education," Science Teacher Education: Vantage Point 1976, 1977 AETS Yearbook. Association for the Education of Teachers in Science and ERIC, 1977.

4. Glass, Gene V.; McGaw, Barry; White, Karl R.: and Smith, Mary Lee. Integration of Research Studies: Meta-analysis of Research. Laboratory of Educational Research University of Colorado, August 1980.

5. Hurd, Paul Dehart. New Directions in Teaching Secondary School Science. Rand McNally and Co., 1969.

6. Kagan, Martin, and Lindman, Harold. "The Proper Experimental Unit Once Again," Journal of Research in Science Teaching. Vol. 17, No. 4, 1980, pp. 351-358.

7. Koran, John J. Jr. "Application of Science Education Research to the Classroom: Validating Science Teacher Training Through Student Performance," Journal of Research in Science Teaching. Vol. 14, No. 1, 1977, pp. 89-86.

8. Kyle, William C. "The Distinction Between Inquiry and Scientific Inquiry and Why High School Students Should be Cognizant of the Distinction," Journal of Research in Science Teaching. Vol. 7, No. 4, 1980, pp. 123-130.

9. Ladd, George T., and Anderson, Hans O. "Determining the Level of Inquiry in Teachers' Questions," Journal of Research in Science Teaching. Vol. 7, No. 4, 1970, pp. 395-400.

10. Morrisey, J. T. "An Analysis of Studies on Changing the Attitude Elementary Student Teachers Toward Science and Science Teaching," Science Education. Vol. 65, No. 2, April 1981.

11. Munby, H. "An Evaluation of Instruments Which Measure Attitudes to Science," World Trends in Science Education. Halifax, Nova Scotia: Atlantic Institute of Education, 1980.

12. Rowe, Marry Budd. "In Pursuit of Quality: Research on the Education of Teachers," Science Teacher Education: Vantage Point 1976, 1977 AETS Yearbook. Association for the Education of Teachers in Science and ERIC, January 1977.

13. Helgeson, Stanley; Howe, Robert W.; and White, Arthur R. Science Education, (Abstracts and index from Research in Education). Education Associates, Inc., Worthington, Ohio, 1966-1972.

14. _____. A Bibliography of Abstracts from Resources in Education, Science, Mathematics, and Environmental Education Information Reference Center, Columbus, Ohio, 1973-1975.

15. _____. A Bibliography of Abstracts from Resources in Education (RIE), Science, Mathematics, and Environmental Education Information Reference Center, Columbus, Ohio, 1976-1977.

16. Welch, Wayne W.; Klopfer, Leopold, E; Aikenhead, Glen S.; Robinson, James T. "The Role of Inquiry in Science Education: Analysis and Recommendations," Science Education. Vol. 65, No. 1, January 1981, pp. 33-50.